

Franklin Quad Core Update/Differences

Helen He

NERSC User Services

yhe@lbl.gov

NERSC User Group Meeting

October 2-3, 2008



Outline

- **This talk is focused on Franklin quad core upgrade and the differences between running on quad core and dual core nodes.**
- **Quad core upgrade plan**
- **Dual and quad core node differences**
- **Compiling and running**
- **Benchmark performance differences**

Current Status

- Franklin is in the middle of its quad core upgrade process, scheduled from July to October 2008.
- Currently Franklin has 4,588 quad core nodes and 1,020 dual core nodes available to the users.
- Quad core environment is the default.
- Franklin "production environment" is a mixture of dual and quad core nodes.
- Franklin quad core nodes are charged as the rate for the dual core nodes. (basically getting 50% discount)
- Upon upgrade completion, Franklin will have 9,660 quad core nodes.

Upgrade Plan

- Done in multiple phases in order to have maximum system availability and job throughput.
- Deliver $\geq 75\%$ of the original computing power on the “production” system throughout the upgrade.
- Limit to 4 upgrade phases to minimize resource commitment and interruption for users.
- Reduce risk of problems by gradually increasing number of upgrade columns during each phase.
- 7-day full system production stabilization time between phases.
- Burn-in time (check out for failed nodes) for the upgraded modules and friendly user time on the “test” system.









Upgrade Schedule

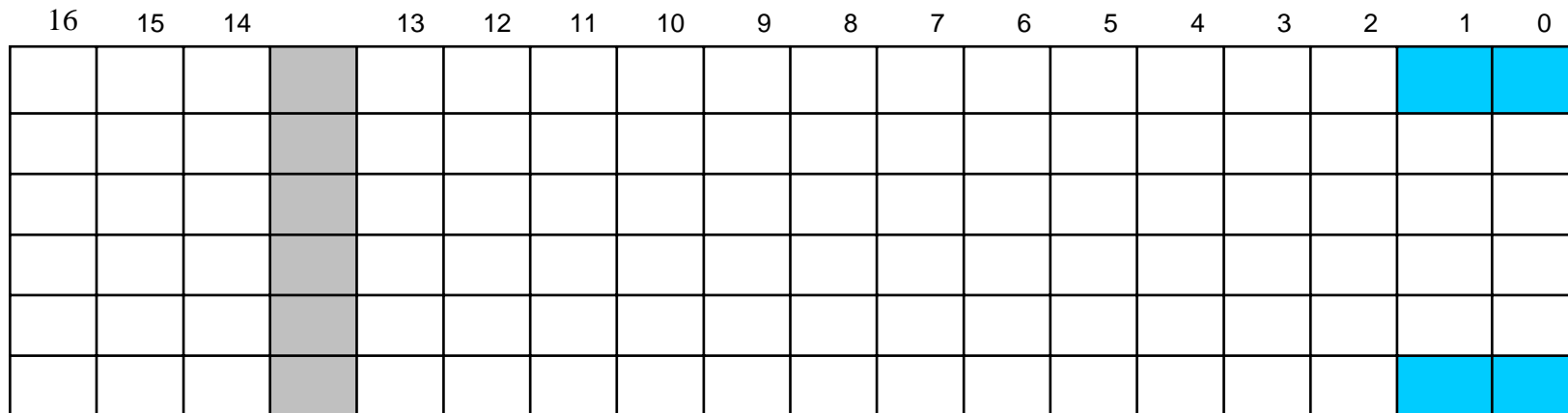
-- Original Configuration

Illustration Scheme from Dan Unger, Cray

- **17 columns x 6 rows = 102 cabinets**
- **Each cabinet has 3 cages * 8 blades * 4 dual core nodes * 2 cores/node = 192 cores.**
- **Total compute cores: 19,320.**

Legend

Empty Column	
Production SIO Module Installed DC	
Production SIO Module Installed QC	
Test SIO Module Installed QC	
Production DC Non-Upgraded	
Test DC Non-Upgraded	
Production QC Upgraded	
Test QC Upgraded	











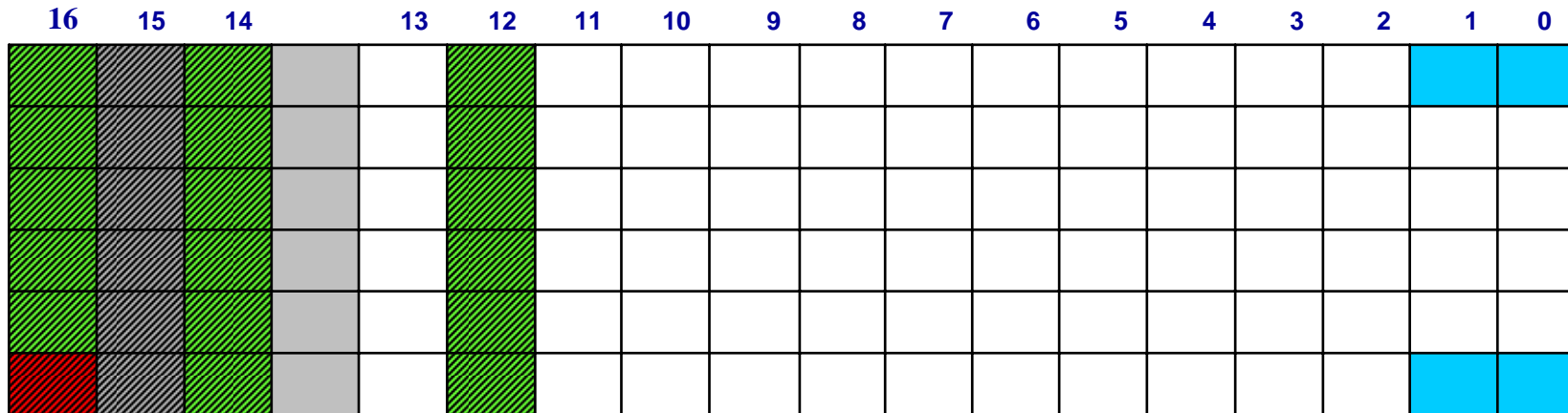
Upgrade Schedule

-- Phase 1 (July 15 - Aug 12)

- **Production: cols 0-11 dual core.**
- **Total compute cores: 14,712.**
- **Franklin is a pure dual core system.**
- **Test system: cols 12,14,16 quad core, 15 dual core, total of 8,032 cores.**

Legend

Empty Column	
Production SIO Module Installed DC	
Production SIO Module Installed QC	
Test SIO Module Installed QC	
Production DC Non-Upgraded	
Test DC Non-Upgraded	
Production QC Upgraded	
Test QC Upgraded	











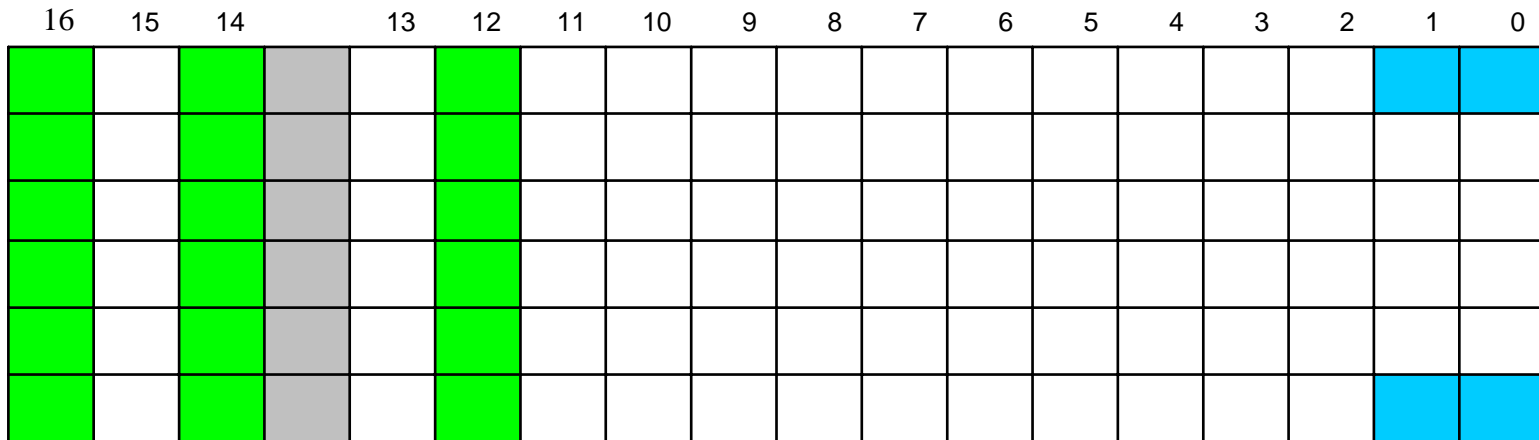
Upgrade Schedule

-- Phase 2a (Aug 13 – Aug 20)

- **Production: cols 12,14,16 quad core, cols 0-10,11,13,15 dual core**
- **Total compute cores: 22,776.**
- **Franklin is a mixed dual core and quad core system.**
- **Dual core environment by default.**

Legend

Empty Column	
Production SIO Module Installed DC	
Production SIO Module Installed QC	
Test SIO Module Installed QC	
Production DC Non-Upgraded	
Test DC Non-Upgraded	
Production QC Upgraded	
Test QC Upgraded	



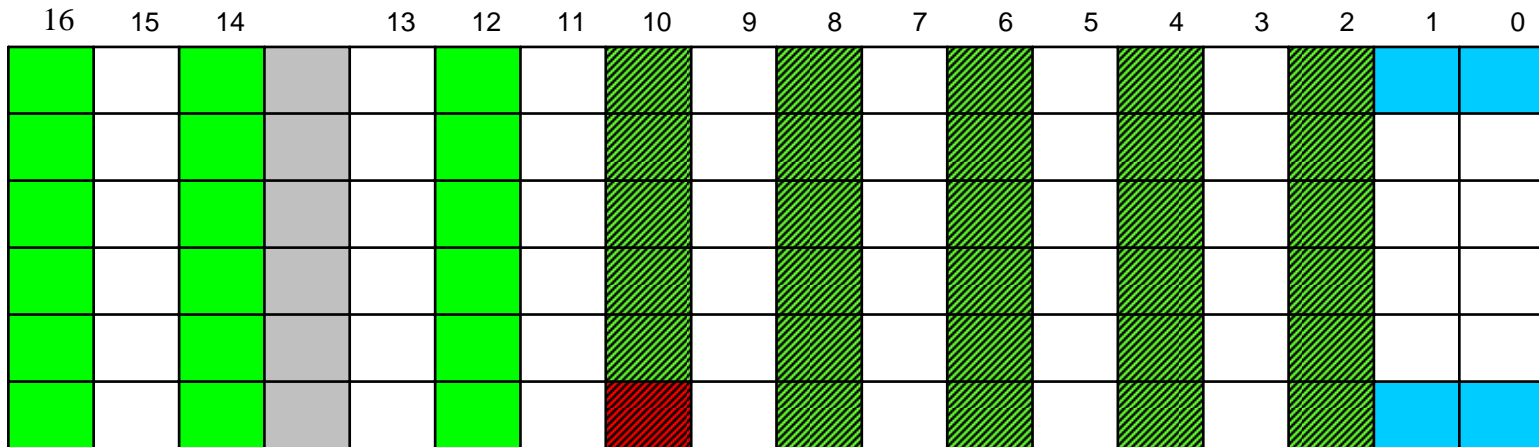
Upgrade Schedule

-- Phase 2b (Aug 21 – Sept 9)

- **Production: cols 12,14,16 quad core, cols 0,1-15 odd dual core**
- **Total compute cores: 17,016.**
- **Franklin is a mixed dual core and quad core system.**
- **Dual core environment by default.**
- **Test system: cols 2-10 even quad core, total of 11,424 cores.**

Legend

Empty Column	Grey
Production SIO Module Installed DC	Light Blue
Production SIO Module Installed QC	Red
Test SIO Module Installed QC	Red with diagonal lines
Production DC Non-Upgraded	White
Test DC Non-Upgraded	Grey with diagonal lines
Production QC Upgraded	Bright Green
Test QC Upgraded	Green with diagonal lines











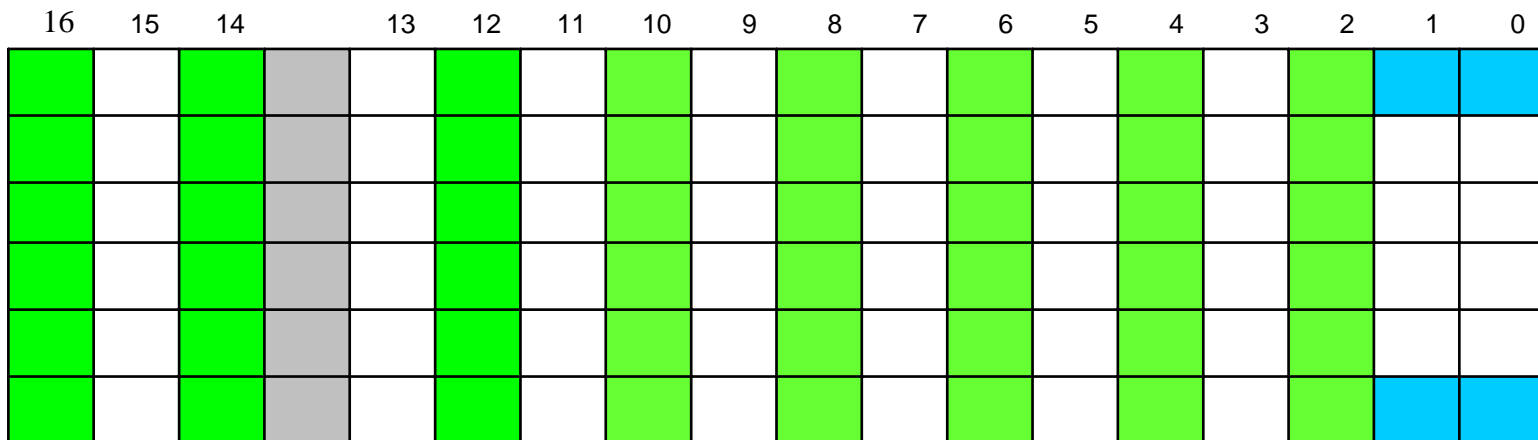
Upgrade Schedule

-- Phase 3a (Sept 10 – Sept 17)

- **Production: cols 2-16 even quad core, cols 0,1-15 odd dual core.**
- **Total compute cores: 28,456.**
- **Franklin is a mixed dual core and quad core system.**
- **Quad core environment by default.**

Legend

Empty Column	
Production SIO Module Installed DC	
Production SIO Module Installed QC	
Test SIO Module Installed QC	
Production DC Non-Upgraded	
Test DC Non-Upgraded	
Production QC Upgraded	
Test QC Upgraded	











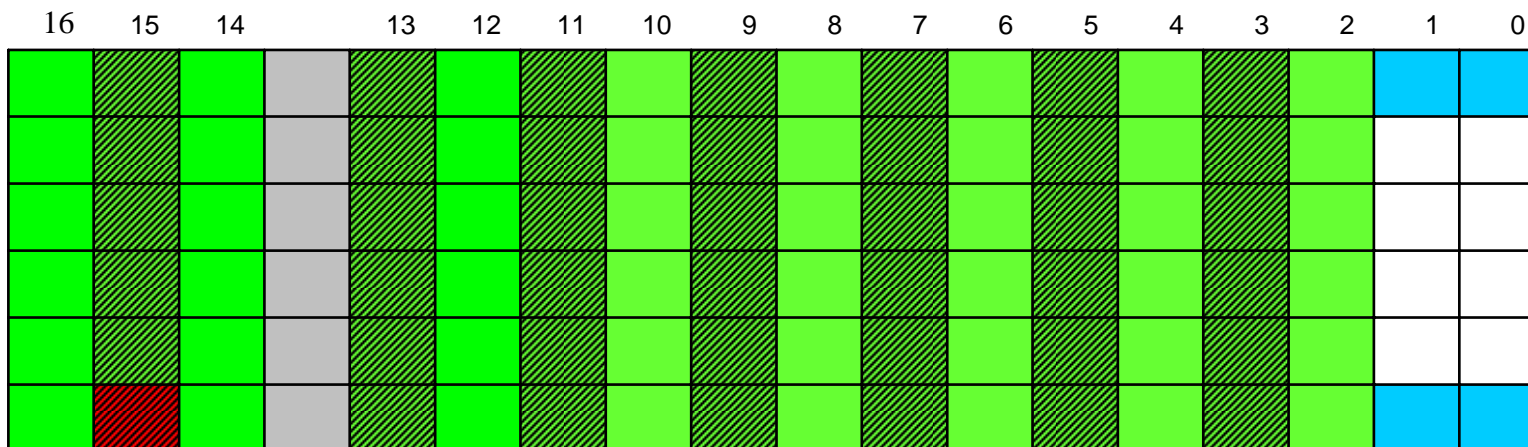
Upgrade Schedule

-- Phase 3b (Sept 17 – ~Oct 6)

- **Production:** cols 2-16 even quad core, cols 0 and 1 dual core
- **Total compute cores: 20,392.**
- **Franklin is a mixed dual core and quad core system.**
- **Quad core environment by default.**
- **Test system:** cols 3-15 odd quad core, total of 16,128 cores.

Legend

Empty Column	
Production SIO Module Installed DC	
Production SIO Module Installed QC	
Test SIO Module Installed QC	
Production DC Non-Upgraded	
Test DC Non-Upgraded	
Production QC Upgraded	
Test QC Upgraded	











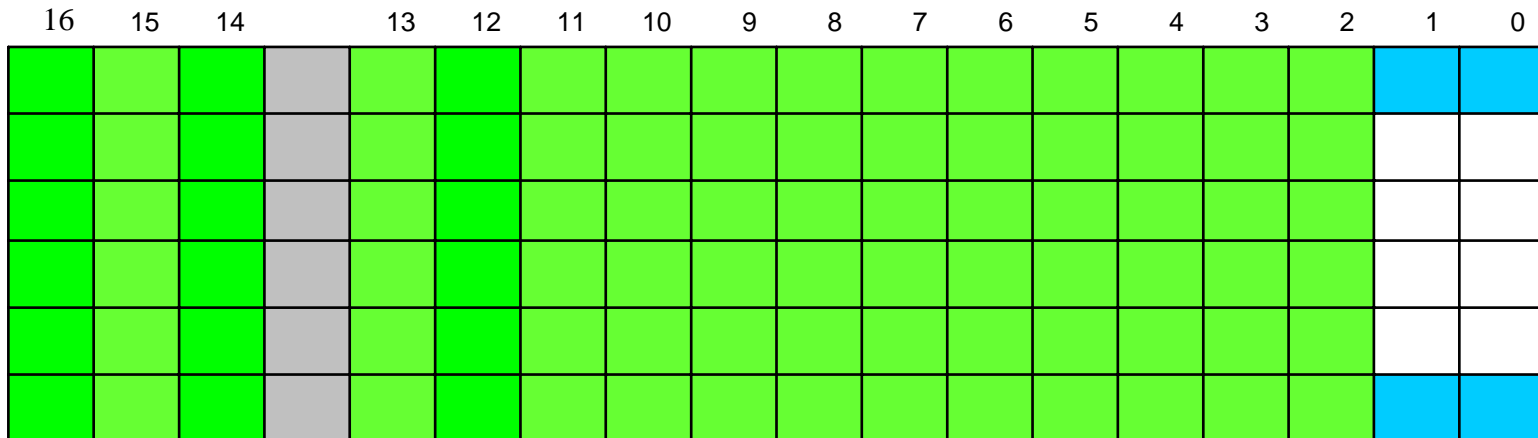
Upgrade Schedule

-- Phase 4a (~Oct 7 – ~Oct 13)

- **Production: cols 2-16 quad core, cols 0 and 1 dual core.**
- **Total compute cores: 36,600.**
- **Franklin is a mixed dual core and quad core system.**
- **Quad core environment by default.**

Legend

Empty Column	
Production SIO Module Installed DC	
Production SIO Module Installed QC	
Test SIO Module Installed QC	
Production DC Non-Upgraded	
Test DC Non-Upgraded	
Production QC Upgraded	
Test QC Upgraded	











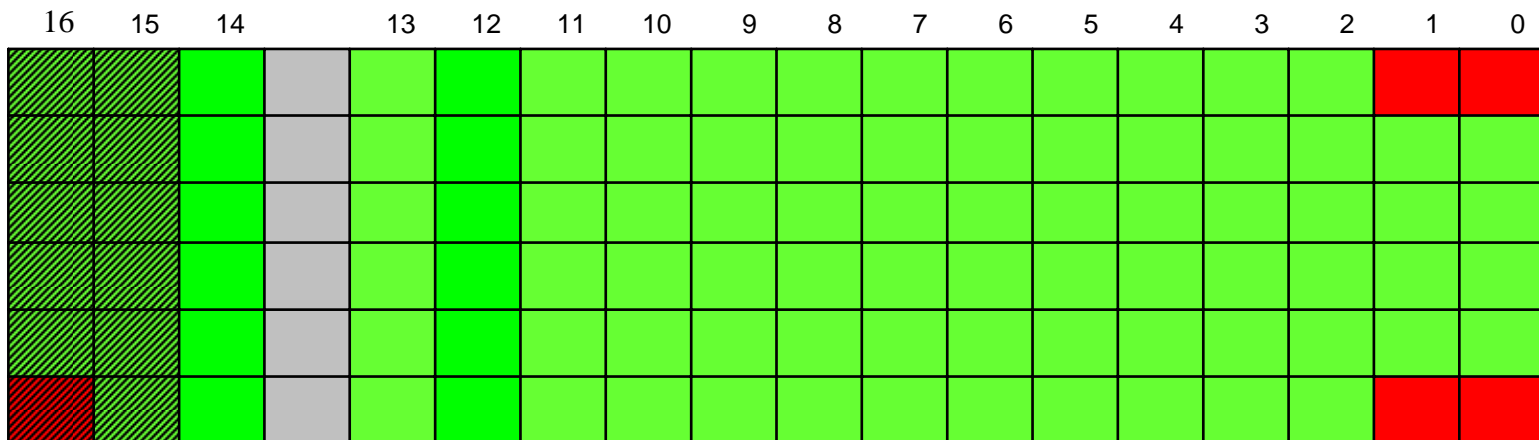
Upgrade Schedule

-- Phase 4b (~Oct 13 – ~Oct 20)

- Swap compute modules in cols 0 and 1 with cols 15 and 16.
- Production: cols 0-14 quad core.
- Total compute cores: 36,600.
- Franklin is a pure quad core system.
- Test system: cols 15 and 16 quad core, total of 4,576 cores.

Legend

Empty Column	
Production SIO Module Installed DC	
Production SIO Module Installed QC	
Test SIO Module Installed QC	
Production DC Non-Upgraded	
Test DC Non-Upgraded	
Production QC Upgraded	
Test QC Upgraded	











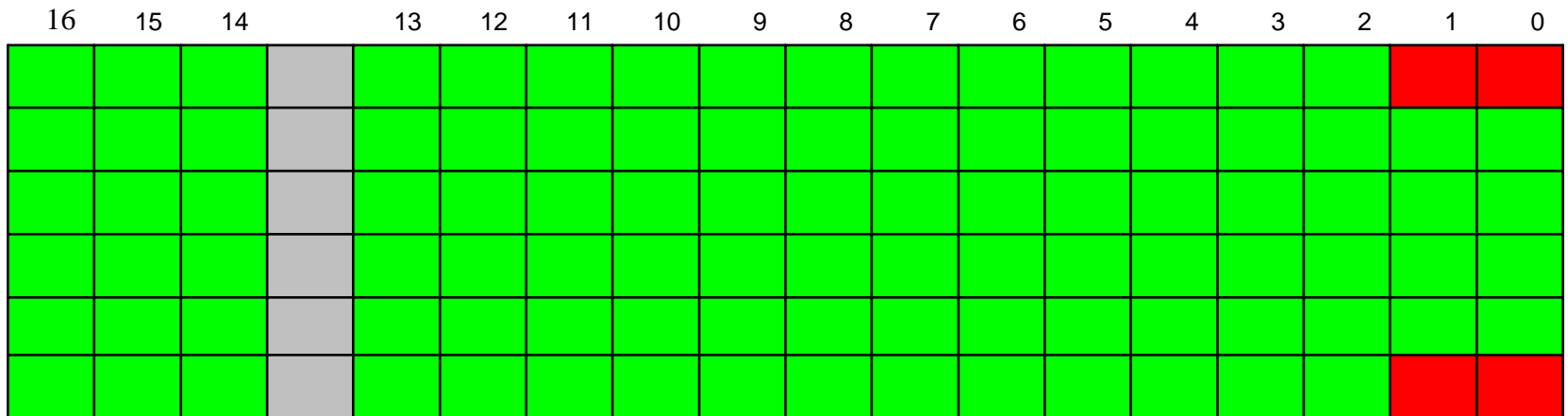
Upgrade Schedule

-- Final Configuration (~Oct 21)

- Phase 4a start date is TBD.
- Phases 4a and 4b may be combined.
- Production: cols 0-16 quad core.
- Total compute cores: 38,640.
- Franklin is a pure quad core system.

Legend

Empty Column	
Production SIO Module Installed DC	
Production SIO Module Installed QC	
Test SIO Module Installed QC	
Production DC Non-Upgraded	
Test DC Non-Upgraded	
Production QC Upgraded	
Test QC Upgraded	



Node Differences

Dual Core Node

- 2.6 GHz clock rate
- 2 flops/cycle = 5.2 GF peak
- Franklin theoretical peak performance: 101.5 TF

- L1 cache: 64 KB/core
- L2 cache: 1 M/core

- Memory: 2 GB/core, 4 GB/node
- Memory speed: 667 MHz

Quad Core Node

- 2.3 GHz clock rate
- 4 flops/cycle = 9.2 GF peak if use SSE128
- Franklin theoretical peak performance: 355 TF

- L1 cache: 64 KB/core
- L2 cache: 512 KB/core
- L3 Shared cache 2 MB/Socket

- Memory: 2 GB/core, 8 GB/node
- Memory speed: 800 MHz

Compiling

- Franklin default environment is quad core.
- Module “xtpe-quadcore” is loaded by default.
- Compiler wrappers include quad core specific compiler options (such as “-tp barcelona-64” for pgi and “-march=barcelona” for pathscale and gnu), and link to quad core Cray LibSci by default.
- Executables built in quad core default environment are targeted to run on quad core nodes, and will not run on dual core nodes (get segmentation fault).
- Codes built in dual core default environment will run on quad core nodes, but probably at lower performance.
- **Strongly recommend to recompile your codes for quad core!**

Compiling (cont'd)

- **To compile for the quad core nodes:**
 - % ftn ...
 - or % cc ...
 - or % CC ...
- **To compile for the dual core nodes:**
 - % module unload xtpe-quadcore
 - % ftn ...
 - or % cc ...
 - or % CC ...

Running

- A single job can be submitted to run on either dual or quad core nodes, but not on a mixture of dual core and quad core nodes.
- #PBS -l feature=quad is set to default.
- #PBS -l mppnppn=4 is set to default.

Running (cont'd)

On quad core nodes:

```
#PBS -q debug  
#PBS -l mppwidth=8  
#PBS -l feature=quad (optional)  
#PBS -l mppnppn=4 (optional)  
#PBS -l walltime=00:10:00  
#PBS -j eo  
cd $PBS_O_WORKDIR  
aprun -n 8 -N 4 ./a.out
```

On dual core nodes:

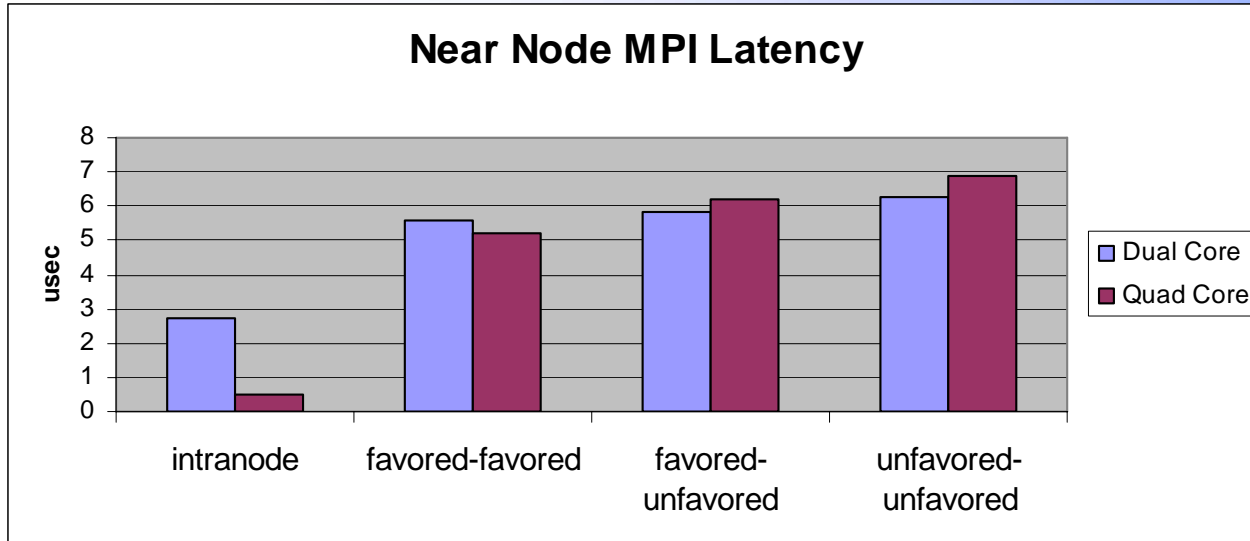
```
#PBS -q debug  
#PBS -l mppwidth=8  
#PBS -l feature=dual (required)  
#PBS -l mppnppn=2 (required)  
#PBS -l walltime=00:10:00  
#PBS -j eo  
cd $PBS_O_WORKDIR  
aprun -n 8 -N 2 ./a.out
```

OpenMP

- With quad core nodes, mixed MPI/OpenMP tests are encouraged.
- The sample script uses 8 nodes, 1 MPI task per node, and 4 OpenMP threads per MPI task.

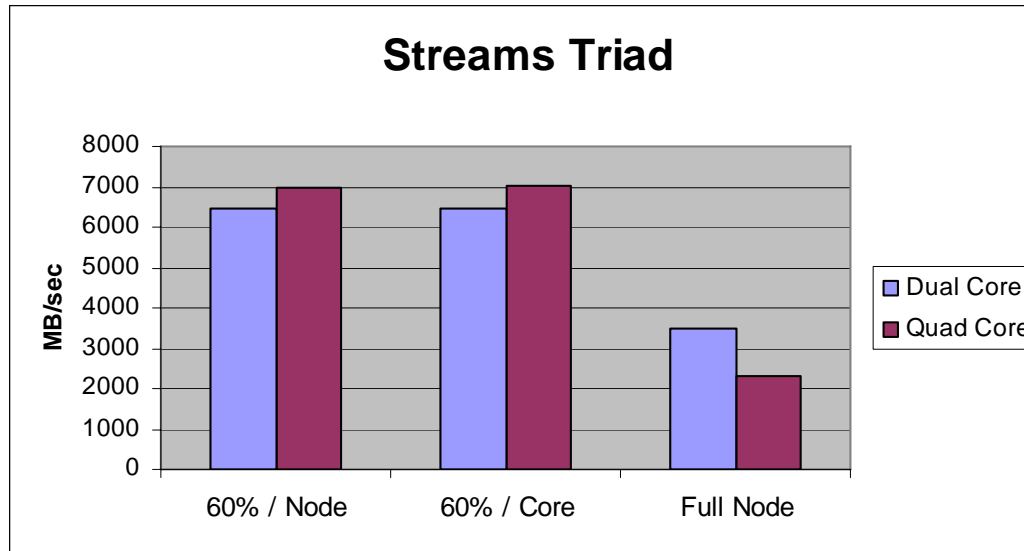
```
#PBS -q debug
#PBS -l feature=quad
#PBS -l mppwidth=8
#PBS -l mppnppn=1
#PBS -l walltime=00:10:00
#PBS -j eo
cd $PBS_O_WORKDIR
ftn -o jac -mp=nonuma jac-openmp.f
setenv OMP_NUM_THREADS 4
time aprun -n 8 -N 1 ./jac
```

MPI Latency



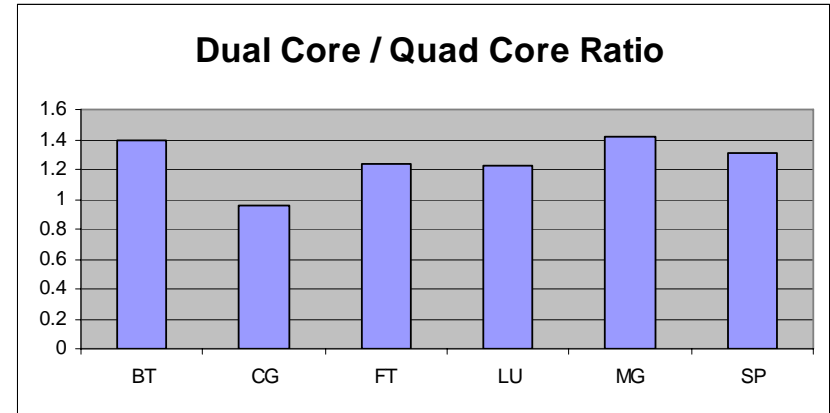
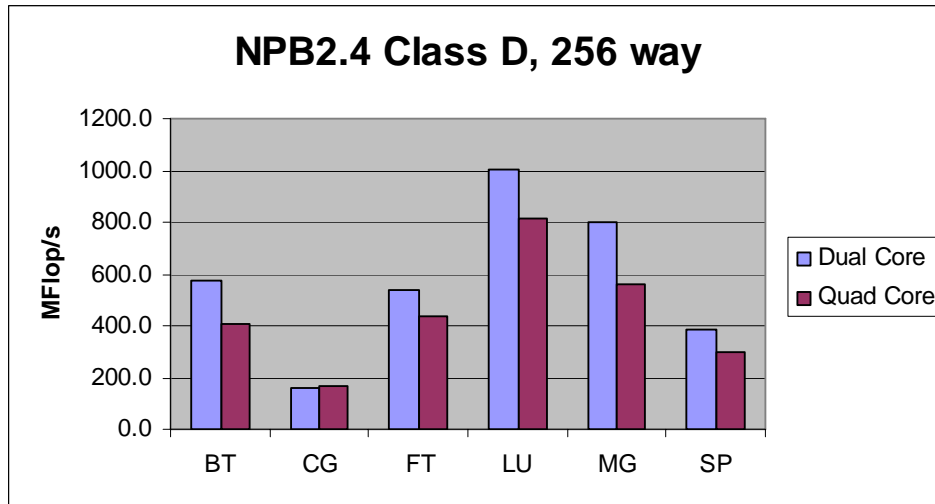
- **One favored core per node.**
 - 1 unfavored core for a dual core node
 - 3 unfavored core for a quad core node.
- **Intranode improvement mainly from xt-mpt/3.**
- **Far node latency is about 1.9 us extra with the 3-D torus Franklin full configuration.**
- **35 hops (between farthest nodes) * 0.053 us (per hop latency) = 1.855, rounded up to 1.9.**

STREAM Benchmark Performance



- **Measures sustained memory bandwidth**
- **Quad core higher:**
 - Single core, use 60% node memory
 - Single core, use 60% core memory
- **Quad core lower:**
 - All cores, use 60% node memory

NPB Benchmark Performance

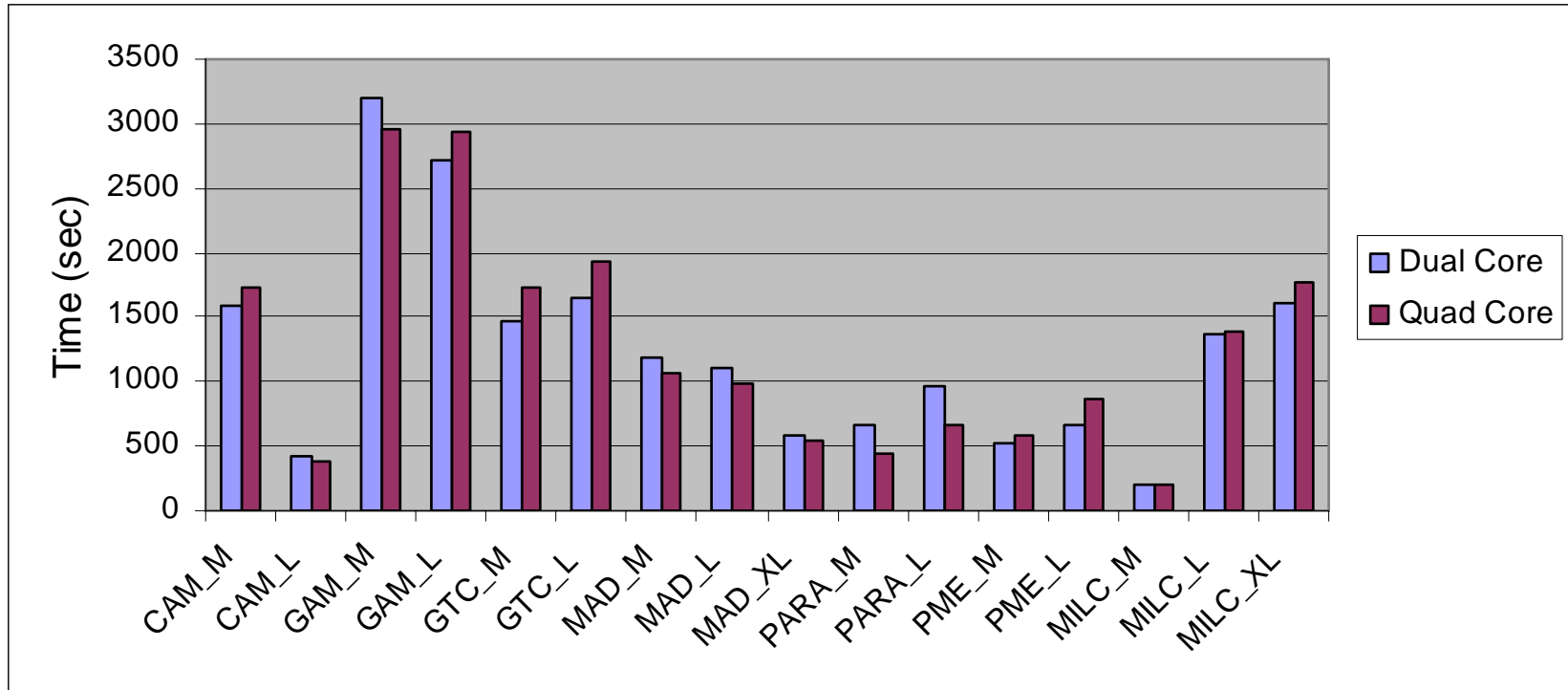


- **NAS Parallel Benchmarks (NPB)**
 - Serial: NPB 2.3 Class B
 - Parallel: NPB 2.4 Class D at 64 and 256 procs
- Quad core mostly slower except for CG.
- Dual core version highly tuned.

Application Benchmarks Summary

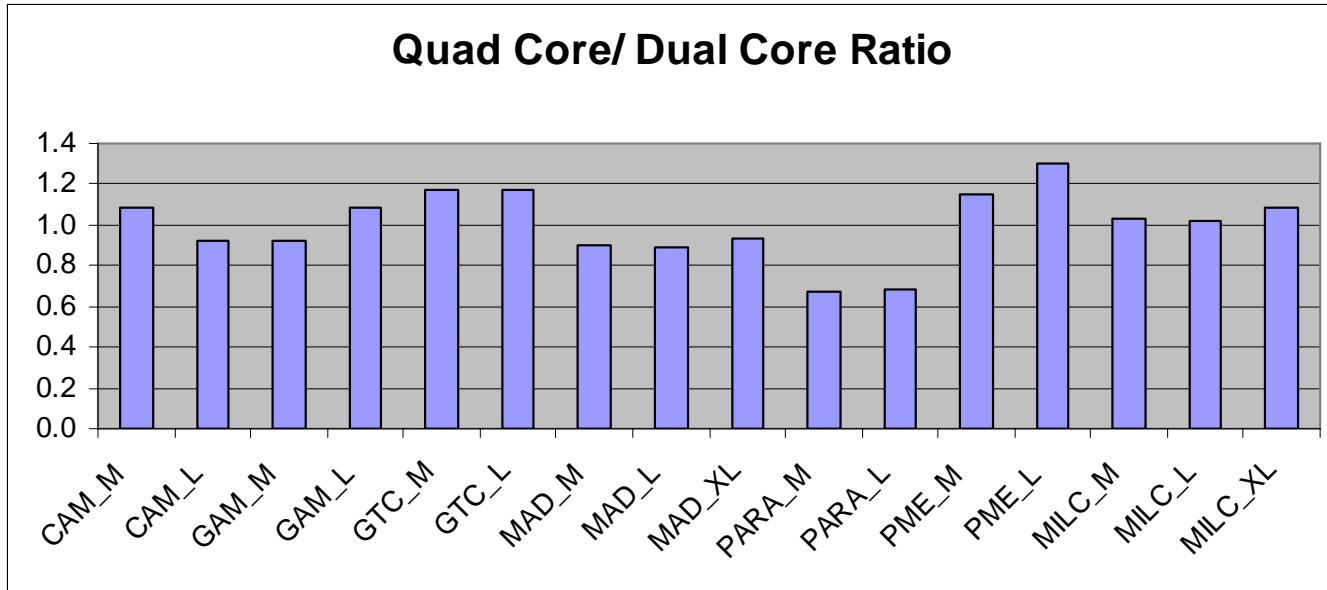
Application	Science Area	Basic Algorithm	Language	Library Use	Comment
CAM3	Climate (BER)	CFD, FFT	FORTRAN 90	netCDF	IPCC
GAMESS	Chemistry (BES)	DFT	FORTRAN 90	DDI, BLAS	DOD TI-06 collaboration
GTC	Fusion (FES)	Particle-in-cell	FORTRAN 90	FFT(opt)	ITER emphasis
MADbench	Astrophysics (HEP & NP)	Power Spectrum Estimation	C	Scalapack	1024 proc. 730 MB per task, 200 GB disk
MILC	QCD (NP)	Conjugate gradient	C	none	2048 proc. 540 MB per task
PARATEC	Materials (BES)	3D FFT	FORTRAN 90	Scalapack	Nanoscience emphasis
PMEMD	Life Science (BER)	Particle Mesh Ewald	FORTRAN 90	none	

Application Benchmarks Performance



- **Medium: 64 cores, except CAM_M 56 cores.**
- **Large: 256 cores, except GAMESS_large 384 cores.**
- **XL: Madbench 1,024 cores, MILC 2,048 cores.**

Application Benchmarks Performance (cont'd)



- Some applications are faster (Madbench, PARATEC), some are slower (GTC, PMEMD).
- Most applications differ within 20% except PMEMD_large.
- PARATEC: >35% faster. Taking advantage of SSE128 optimization.
- MILC: quad core extra tuning.

A Few Misc Topics

- CPU clock rate reduced, but memory speed improved.
- Overall application performances (NERSC Sustained System Performance) are about the same (~1% difference).
- Quad core nodes have the same charging factor as dual core nodes, but are charged only 2 cores/node for the allocation year 2008.
- The min number of cores used for reg_big and reg_xbig queues (to get 50% discount) are doubled. Min number of nodes stay the same.
- Due to the reduced number of nodes available, average queue wait time is longer. Please be patient, the situation will become better after upgrade completes.
- Before Franklin quad core is officially accepted, performance results could now be published AFTER consulting with NERSC.

More information

- **Franklin web page:**
<https://www.nersc.gov/nusers/systems/franklin/>
- **Franklin Quad Core Upgrade Plan web page:**
http://www.nersc.gov/nusers/systems/franklin/quad_core_upgrade.php